

FLOATING-POINT PROCESSOR WITH OPERATING MODE HAVING IMPROVED ACCURACY AND HIGH PERFORMANCE

ABSTRACT OF THE DISCLOSURE

5 Floating-point units (FPUs) and processors having a “flush-to-nearest”
operating mode that provides improved accuracy over a conventional “flush-to-zero” mode.
The FPU or processor includes an operand processing section and an operand flush section.
For each floating-point operation, the operand processing section receives and processes one
or more input operands to provide a preliminary result. The operand flush section determines
10 whether the preliminary result falls within one of a number of ranges of values and sets the
preliminary result to one of a number of set values if the preliminary result falls within one of
the ranges. In a specific implementation, a first range of values is defined to include values
greater than zero and less than half of a minimum normalized number (i.e., $0 < |y| < +a_{\min}/2$),
a second range of values is defined to include values equal to or greater than $+a_{\min}/2$ and less
15 than $+a_{\min}$ (i.e., $a_{\min}/2 \leq |y| < a_{\min}$), and the preliminary result is set to zero if it falls within the
first range and to $+a_{\min}$ or $-a_{\min}$ (depending on the sign bit) if it falls within the second range.